G. SCOTT EMBLIDGE, State Bar No. 121613
emblidge@meqlaw.com
RACHEL J. SATER, State Bar No. 147976
sater@meqlaw.com
ANDREW E. SWEET, State Bar No. 160870
sweet@meqlaw.com
MOSCONE, EMBLIDGE, & QUADRA, LLP
220 Montgomery Street, Suite 2100
San Francisco, California 94104-4238
Telephone: (415) 362-3599
Facsimile: (415) 362-2006

Attorneys for Plaintiff

UNITED STATES DISTRICT COURT FOR THE

NORTHERN DISTRICT OF CALIFORNIA

COYNESS L. ENNIX JR., M.D., Plaintiff, vs. ALTA BATES SUMMIT MEDICAL CENTER, Defendants.	Case No.: C 07-2486 WHA **DECLARATION OF WILLIAM S. WEINTRAUB, M.D., IN OPPPOSITION TO DEFENDANT'S MOTION FOR SUMMARY JUDGMENT** **Date: April 24, 2008** **Time: 8:00 a.m.** **Dept: Ctrm. 9, 19th Floor** **Judge: Hon. William H. Alsup**

I, William S. Weintraub, M.D., declare:

1. I am a Cardiology Section Chief at Christiana Care Health System and the Director of the Christiana Care Center for Outcomes Research. I have personal knowledge of the facts stated in this declaration and Exhibits A and B.

2. I have been an active clinician and educator. I am the director of the Cardiovascular Fellowship Program at Christiana. As a scientist, I am a cardiovascular

epidemiologist. I have been an active investigator for 30 years, and have over 340 peer reviewed publications. As an epidemiologist and active cardiovascular investigator, I have expertise in the assessment of risk. I have written extensively on outcomes of revascularization procedures. Among other things, I was one of the founders of the American College of Cardiology National Cardiovascular Registry (NCDR). This registry includes more than 2 million case records, making it the largest clinical database examining outcomes in cardiology in the world. I remain on the NCDR management board and chair the Cath-PCI steering committee. As the head of cardiology for a large health care system, I am actively involved in the credentialing process for cardiologists.

3. I reviewed the documents relating to the use of statistics in the peer review process regarding Dr. Coyness Ennix. Attached to this declaration as Exhibits A and B are true and correct copies of the two reports I provided in this litigation on that subject. The statements in those declarations are true and correct.

4. I concluded that the totality of the evidence reveals that Dr. Ennix cannot be shown to be meaningfully different in practice from his peers in any way. In point of fact the very data on which the Medical Staff relied in criticizing Dr. Ennix could be used to establish the proficiency of Dr. Ennix as a cardio-thoracic surgeon.

5. I also reviewed the statistical approach used by the State of California in its report concerning mortality rates in 2003 and 2004 for cardiac surgeons statewide performing isolated coronary bypass surgeries, also known as the CCORP report. The approach taken in the CCORP report is consistent with the best mathematical methods for risk adjustment, using the multivariable approach logistic regression to correct for confounding. In this report Dr. Ennix is not listed as a surgeon with worse performance than peers. His risk adjusted mortality for coronary bypass surgery straddles the state average, meaning that there is no evidence that Dr. Ennix's performance, based on mortality is worse than the state average of his peers. This report should be considered the definitive assessment of surgical mortality for isolated coronary bypass surgery by surgeon for the State of California. There are two reasons for this, first related to the

large sample size of the control number of surgeons and surgeries (that is all other providers), and the second is that the State of California report is created independently of any one health care system.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct and that this declaration was signed in Newark, Delaware.

Dated: March __, 2008

/s/
William S. Weintraub, M.D.

EXHIBIT A

03/27/2008

Assessment of Outcomes of Coyness Ennix, MD

William S. Weintraub, MD, FACC, FAHA
Professor of Medicine Jefferson University
Professor of Medicine, Emeritus, Emory University
Professor of Public Health, Emeritus, Emory University
Professor Health Sciences, Adjunct, University of Delaware
Cardiology Section Chief, Christiana Care Health System
Director, Christiana Care Center for Outcomes Research

I am currently Cardiology Section Chief at Christiana Care Health System in Newark, Delaware. I am also the Director of the Christiana Care Center for Outcomes Research. I graduated from Johns Hopkins University with an MD degree in 1975. I trained at Boston University, The Mount Sinai School of Medicine and the University of Pennsylvania. I am board certified in Internal Medicine and Cardiovascular Diseases. My first faculty appointment was at the University of Pennsylvania. I have been on the faculty of Emory University since 1986. I am currently Professor Emeritus of Medicine and Professor Emeritus of Public Health. My appointments in Public Health have been in Health Policy and Management and Epidemiology. My current academic appointments are as Professor of Medicine at Thomas Jefferson University and Professor of Health Sciences at the University of Delaware. I have been an active clinician and educator. I am the director of the Cardiovascular Fellowship Program at Christiana. As a scientist, I am a cardiovascular epidemiologist. I have been an active investigator for 30 years, and have over 340 peer reviewed publications. As an epidemiologist and active cardiovascular investigator, I have expertise in the assessment of risk. I have written extensively on outcomes of revascularization procedures. Among other things, I was one of the founders of the American College of Cardiology National Cardiovascular Registry (NCDR). This registry includes more than 2 million case records, making it the largest clinical database examining outcomes in cardiology in the world. I remain on the NCDR management board and chair the Cath-PCI steering committee. As the head of

cardiology for a large health care system, I am actively involved in the credentialing process for cardiologists.

Epidemiology stems from the Greek meaning study of epidemics, but is fundamentally the study of risk. The first great epidemiologic study by John Snow is a classic example. Snow showed that during an episode of cholera in the 19th century, all cases were in people drawing water from a single pumping station. This finding had a tremendous impact on public health, bringing forth the importance of clean water. The same principles of evaluating risk and using this evaluation to guide public policy remains the purpose of epidemiology to this day.

Risk is assessed by finding a higher incidence of an adverse outcome associated with some particular risk factor. Risk is the probability or incidence rate of an outcome occurring. A related idea is the odds, the ratio of an event occurring to it not occurring. The impact of a risk factor may then be assessed by the relative risk, the ratio of risk of an adverse outcome with the risk factor to the risk without the risk factor. Similarly the odds ratio is the ratio of the odds with the risk factor to the odds without the risk factor. For each of these measures, there is no increase in risk with the risk factor if the relative risk or the odds ratio is one or less, while if greater than 1 there is increased risk. However, there is uncertainty in the measurement of the relative risk or odds ratio. This uncertainly is related to the stochastic or probabilistic nature of measurement. Thus, these ratios are associated with confidence intervals. Most commonly the 95% confidence interval is used. This interval means that if results are sampled, the measurement would fall within this boundary 95% of the time. For risk to be increased, the 95% confidence interval for the relative risk or odds ratio must be greater than one. Thus, an odds ratio of 1.5 with 95% confidence interval of 1.3 to 1.8 means that the odds are increased by 50%, with a 95% chance that the measurement lies between 30 and 80% increased. The 95% confidence interval is dependent on sample size. Thus, for a sample of 1000, the 95% confidence interval will be much smaller than for a sample size of 100. The stochastic nature of measurement of risk also means that if multiple risk factors are considered, then using the 95% confidence interval or the p value (the probability of

falsely finding an association when there isn't one) of less than 0.05, means that on average one of risk factors will be falsely positive for every twenty considered.

Measurement of increased risk does not mean that the risk factor under consideration has caused the adverse outcome. This is because the measurement of increase risk may be confounded. A confounder is a risk factor that is associated with an adverse outcome, and is also associated with a putative risk factor under consideration. For instance, studies have shown that earlobe creases are associated with increased risk of cardiovascular events. However, cigarette smoking causes wrinkling of the skin and is generally agreed to be causative of cardiovascular events. If cigarette smoking causes both the cardiovascular disease as well as the earlobe creases, then the earlobe creases are an epiphenomenon, and not causally related to the events. The issue of cigarette smoking and causation let the British epidemiologist Sir Austin Bradford Hill to devise a set of criteria that are often used to assess causation. This criteria include consistency of observation, strength of association, dose response, temporal precedent (the risk factor must occur before the event), biological plausibility, coherence (causation is not contradicted by other facts), and experimental evidence. To decide that a putative risk factor is causative, it should be consistent with the Bradford Hill criteria, and then be generally agreed to. Thus, over time cigarette smoking has come to be viewed as meeting the Bradford Hill criteria for causation of cardiovascular events.

Experimental design in epidemiologic studies tries to minimize bias while assuring adequate sample size to achieve a small 95% confidence interval. Investigators use three different type of studies evaluate risk, randomized trials, cohort studies and case control studies. Randomized trials are most useful when comparing a new form of therapy to placebo or to the previous standard. Randomization, in principle, eliminates selection bias, that is it eliminates confounding. Cohort studies and case control studies are generally used to assess risk factors where exposure cannot be randomized. Cohort studies involve tracking subjects over time and following outcome. In case control studies, investigators work backwards, starting with an outcome and then establishing whether those with the outcome are more likely to have the risk factor of interest. Case

control studies are most useful when studying rare outcomes, but relatively common risk factors. Both cohort studies and case control studies may suffer from confounding. Investigators try to eliminate confounding by using mathematical techniques, such as the multivariable methods logistic regression or Cox model analysis. This can work adequately when the confounding variables can be measured, but no mathematical technique can eliminate unmeasured confounders.

The problem of assessing outcomes of hospitals or health care providers poses many challenges for the epidemiologist. One of the most challenging is the assessment of surgical outcomes using event rates. Such an analysis is a form of cohort study, in which one provider is compared to peers. Outcomes may fairly be used for peer review, in an educational process, to help all providers and operators to achieve better outcomes. However, to use outcomes, or scorecards, to assess competency, or that the risk of an adverse outcome is higher for one surgeon than for his peers is difficult, and it is unproven that this can be successfully accomplished. Thus, a mathematical measure of outcome to assess performance and establish privileges is at best uncertain. There are several reasons for this, including that the number of procedures performed by any one surgeon will generally be low, creating uncertainty in the measurement, as noted above. There may also be differences in patients operated on by different surgeons, that is confounding. Also, if mortality of multiple surgeons is assessed, with 95% confidence intervals, then on the basis of probability one in 20 would have a 95% confidence interval that did not include unity even if all providers had the same outcome. There will also be difference between operations, with more frequent adverse events with higher risk surgery. For some outcomes, such as blood loss, there may be subjectivity to the measurement, giving rise to another type of bias, misclassification. If one measure, such as transfusion, is use to assess another measure, i.e., blood loss, then the analysis requires yet more stringent statistical approaches. It is also difficult to choose appropriate outcomes to measure. For an outcome to be a meaningful measure, it must occur sufficiently frequently that statistical approaches can be meaningfully measured, there must be potential variation between providers and the outcome must depend on provider performance. Thus, mortality rates for quite rare operations cannot be approached in a

statistically meaningful way. Similarly, mortality for operations on patients in cardiogenic shock is not a meaningful outcome measure, as the expected mortality is quite high and more dependent on how sick the patient is than the performance of the provider. There should also be a limited number of outcomes considered, best agreed to in advance. If too many outcomes are considered, and if new outcomes are considered after the investigation begins, then there is potential for outcomes to be considered until one proves positive, i.e. data dredging. When multiple outcomes are considered, statistical correction, such as proposed by Bonferroni should at least be considered. Finally, in considering using statistics to evaluate the outcome of a health care provider, it is necessary to consider the totality of the evidence, and not cherry pick one particular outcome which seems to be worse than the rest.

The ad hoc committee which considered the performance of Dr. Ennix entered into the daunting task of using statistics to evaluate whether performance was substandard.

In examining the records from Summit, return to surgery is offered without statistical comparison of any type. His return to surgery rates from 2002 through 2005 was 6.4%. This is in comparison to 4.7% by all non-Kaiser providers. His conversion from off pump to cardio-pulmonary bypass was 2%, as a point estimate the lowest of his peers. Note that these statistics are not adjusted for any confounding variable and include all operations. No statistical inference testing was performed. It is not possible to see how these "outcomes" can meaningfully be used to assess performance.

An analysis mortality after CABG surgery by Dr. Richard Shaw used the approach of assessing the observed to expected (O/E) ratio. Similar to the odds ratio, if the O/E ratio is greater that 1 and the 95% confidence interval does not include 1, then there is increased risk. The analysis showed that from the year 2000 until 2005 there was no trend to increased mortality risk, with the O/E ratio compared to other surgeons at times above 1, at times below 1, but always with 95% confidence intervals that included 1. In addition, the confidence intervals with the other surgeons at Summit overlapped, meaning that there was no statistical difference between Dr. Ennix and the other surgeons at

Summit. The data in Dr. Shaw's report comes from the Summit hospital mortality data, which was submitted to the Society of Thoracic Surgeons (STS), for the STS national database. I have reviewed these data as well, and completely concur with Dr. Shaw's report.

The State of California also provides a report on mortality after coronary artery bypass grafting for the years 2003-2004. Note that Dr. Ennix is one of just 3 cardio-thoracic surgeons practicing in the State of California who is on the clinical advisory panel for this report. The statistical approach used in this report is consistent with the best mathematical methods for risk adjustment, using the multivariable approach logistic regression to correct for confounding. The methods section of this report shows that the panel is well aware of both their social responsibility to report outcomes and the methodological difficulties noted above. In this report Dr. Ennix is not listed as a surgeon with worse performance than peers. His risk adjusted mortality for coronary bypass surgery straddles the state average, meaning that there is no evidence that Dr. Ennix's performance, based on mortality is worse than the state average of his peers. This report should be considered the definitive assessment of surgical mortality for isolated coronary bypass surgery by surgeon for the State of California. There are two reasons for this, first related to the large sample size of the control number of surgeons and surgeries (that is all other providers), and the second is that the State of California report is created independently of any one health care system. To independently of this report establish another statistical basis for evaluating outcome for Dr. Ennix that would show worse outcomes than his peers would be very difficult indeed.

It is not possible to look statistically at other operations besides CABG performed by Dr. Ennix or the other surgeons at Summit as the numbers were too low for meaningful statistical analysis. However, as point estimates the number do not seem to vary from his peers.

03/27/2008

Summary

Establishing a statistically significant, meaningful difference in surgical performance based on outcome measures is a challenging task. Using such to establish or deny credentials is not an established practice. In the setting of outliers, where performance is clearly different from peers, it may be possible to use outcome measures for credentialing. This relates to the concept of strength of association, as noted in the Bradford Hill criteria. While such a standard to even pick such outliers is uncertain, any such reasonable standard that one might create of less than acceptable performance has most certainly not been met for Dr. Ennix. Furthermore, such a standard should be created and agreed to in advance, if done at all. The material from Summit shows no such standard. For some measures as point estimates his performance is worse than peers, for some better. Thus, there is no evidence of consistency of association suggesting poor outcome. In any case, for none is there statistical evidence of worse performance. The totality of the evidence reveals that Dr. Ennix cannot be shown to be meaningfully different in practice from his peers in any way. In point of fact these very data could be used to establish the proficiency of Dr. Ennix as a cardio-thoracic surgeon.

I have not been deposed or served as a witness at trial in the last four years. I am paid $600 an hour.

EXHIBIT B

03/27/2008

Response to the report by Gregory Duncan

William S. Weintraub, MD, FACC, FAHA

Professor of Medicine Jefferson University

Professor of Medicine, Emeritus, Emory University

Professor of Public Health, Emeritus, Emory University

Professor Health Sciences, Adjunct, University of Delaware

Cardiology Section Chief, Christiana Care Health System

Director, Christiana Care Center for Outcomes Research

I have reviewed the report by Dr. Gregory Duncan. Dr Duncan developed a small dataset of some 56 rows. There are just some 5 column variables in the dataset. I have not previously seen this dataset and it is not explained at all. Apparently this analysis is derived from the summary tables provided by Alta Bates Medical Center. The first column is whether the surgeon is Dr. Ennix or not, the second column the type of procedure, the third column the number of procedures, the fourth column the number of deaths and the fifth column the year. The rows are grouped by year, with the first four for each year being Dr. Ennix and the second four being other Alta Bates surgeons. In each group of four, the rows are aggregate, isolated coronary surgery, isolated valve and valve plus coronary surgery. If we leave out the aggregate, that leaves just 42 observations. For the aggregate data there would be just 14 observations. Dr. Duncan then uses a logistic regression model to estimate the relative mortality of Dr. Ennix compared to peers. Dr. Duncan found an odds ratio of 2.08 (95% CI 1.34 to 3.25) for the 14 observations in the aggregated data. This is an analysis at the level of summary data rather than an analysis at the level of the individual case. There is no effort to adjust for any standard confounding variable such as age, gender, diabetes, hypertension, acuteness, or repeat procedure. For instance, if Dr. Ennix's patients were at higher risk because of older age or more severe disease, this would not be reflected in Dr. Duncan's analysis. In point of fact, there is an extensive literature on patient level assessment of outcome, and this is the standard in the field, used not only in California, but in New York as well.

Furthermore, to state that there is no heterogeneity between types of procedures, coronary surgery and valve surgery is ludicrous. This just speaks to the analysis being underpowered. It makes no sense to include in one analysis outcomes of valve surgery and coronary surgery; it is well known that the mortality of valve surgery, especially mitral valve surgery, is higher than of coronary surgery. Dr. Duncan also states that there is no interaction by year, but this is not correct. There is a statistically significant interaction by year, with increased risk largely in 2004, the year Dr. Ennix introduced minimally invasive surgery. Dr. Ennix has admitted this did not go well, and has agreed not to do this type of procedure again. Thus, there is a clear explanation of increased risk for this one year. Dr. Duncan also ignores the discrimination of the model. Discrimination is often measured with a c index, where a c index of 0.5 means that there

is no ability to tell who will have an event after the procedure and 1.0 perfect ability to predict. The full model with main effects and interactions only has a c index of 0.702. This contrasts poorly with models for mortality at the patient level for coronary surgery which often have a c index of between 0.75 and 0.9. However, the c index in this case was increased dramatically by the interactions as the main effects model, which compares Dr. Ennix to his peers, adjusted for year and surgery type was only 0.572. This reflects almost no ability to determine who will die after a procedure. This also means that there are other variables, not available in the aggregated dataset, that predict mortality. The higher c index for the full model with interactions just shows that Dr. Ennix had higher mortality in just 2004, which is a special case because of the minimally invasive surgery.

Dr. Duncan also misinterprets my original position that it is difficult to show that one physician is performing worse than another. I have never stated that we could show that Dr. Ennix is no worse, or more analytically non-inferior. We do not purpose that the null hypothesis must be accepted, but rather that the alternative hypothesis that Dr. Ennix has worse outcomes cannot be satisfactorily proven. A non-inferior design would require establishing a margin and upper bound of the 95% confidence interval for what would be considered worse outcome. I doubt that such an approach has ever been attempted for physician performance.

This is quite simply the most irresponsible analysis of physician performance I have seen in over 25 years of experience as an active cardiovascular investigator, clinician, teacher and administrator. Implicit in this analysis is that Dr. Duncan believes that this analysis offers causality, that is that Dr. Ennix is causally responsible for increased mortality. I can say this because if causality cannot be shown, then the analysis is of no value. However, Dr. Duncan offers no argument whatsoever that this analysis offers a causal responsibility on the part of Dr. Ennix for the increased mortality, which has not even been shown to be the case as noted above.

Dr. Duncan ignores the report from the State of California which finds no increased mortality for Dr. Ennix for coronary surgery, where proper case mix adjustment was employed. Needless to say, the analysis by the State of California compared Dr. Ennix to a larger number of peers and used a patient level analysis. It was conducted by a neutral third party as well. In addition, the California model had a main effects c-index

of 0.801, far better than the main effect model offered by Dr. Duncan. To offer this analysis as an alternative to the analysis from the State of California makes no sense.

Dr. Duncan's analysis is basically made up by him, and as far as I can tell is entirely without precedent as to the methods used for evaluating physician performance. Furthermore, Alta Bates commissioned this analysis after my report. Thus, Alta Bates is offering this highly flawed post hoc analysis. Dr. Duncan's analysis is entirely without merit and Alta Bates approach to evaluating physician performance as to quantitative assessment of outcome is irresponsible.

In reviewing Dr. Duncan's qualifications, I note that he has a background in statistics and economics, but I see no evidence of experience in biostatistics, epidemiology, or outcomes assessment. I see no evidence of knowledge about design of studies in clinical medicine. I see no evidence that Dr. Duncan would be familiar with issues related to assessing physician performance. I see no evidence that Dr. Duncan would have knowledge or experience concerning outcomes of cardiothoracic surgery at the level of a biostatistician working with such data. Dr. Duncan is not qualified to evaluate physician performance.

The review and findings of this report are all my own. I have consulted with a senior faculty biostatistician to confirm these findings.